Bias-Correcting the Realized Range-Based Variance in the Presence of Market Microstructure Noise*

Kim Christensen† **Mark Podolskij**‡ **Mathias Vetter**§

This version: October 10, 2006

Abstract

Market microstructure noise is a challenge to high-frequency based estimation of the integrated variance, because the noise accumulates with the sampling frequency. In this paper, we analyze the impact of microstructure noise on the realized range-based variance and propose a bias-correction to the range-statistic. The new estimator is shown to be consistent for the integrated variance and asymptotically mixed Gaussian under simple forms of microstructure noise, and we can select an optimal partition of the high-frequency data in order to minimize its asymptotic conditional variance. The finite sample properties of our estimator are studied with Monte Carlo simulations and we implement it on high-frequency data from TAQ. We find that a bias-corrected range-statistic often has much smaller confidence intervals than the realized variance.

JEL Classification: C10; C22; C80.

Keywords: Bias-Correction; Integrated Variance; Market Microstructure Noise; Realized Range-Based Variance; Realized Variance.

*Parts of this paper were written, while Kim Christensen was at the University of California, San Diego, whose hospitality is gratefully acknowledged. Mark Podolskij and Mathias Vetter were supported by the Deutsche Forschungsgemeinschaft grant SFB 475 "Reduction of Complexity in Multivariate Data Structures." The code for this paper was written in the Ox programming language, due to Doornik (2002).

†Aarhus School of Business, Dept. of Marketing and Statistics, Haslegaardsvej 10, 8210 Aarhus V, Denmark. Phone: (+45) 89 48 63 74, fax: (+45) 86 15 39 88, e-mail: kic@asb.dk.

‡Ruhr University of Bochum, Dept. of Probability and Statistics, Universitätstrasse 150, 44780 Bochum, Germany. Phone: (+49) 234 32 28330, fax: (+49) 234 32 14559, e-mail: podolski@cityweb.de.

§Ruhr University of Bochum, Dept. of Probability and Statistics, Universitätstrasse 150, 44780 Bochum, Germany. Phone: (+49) 234 32 23283, fax: (+49) 234 32 14559, e-mail: mathias.vetter@rub.de

1. Introduction

In this paper, we analyze the impact of noisy high-frequency data on the *realized range-based variance* (RRV) (see, e.g., Parkinson (1980), Christensen & Podolskij (2006*a*, 2006*b*) or Dijk & Martens (2006)). We propose a new robust range-based estimator, which is consistent for the *integrated variance* (IV) and asymptotically mixed Gaussian in the presence of simple forms of microstructure noise. Moreover, we show how to optimally divide the high-frequency data such that the conditional variance of the asymptotic distribution is minimized.

Our paper is motivated by the increasing use of high-frequency data to measure the ex-post variation of asset price processes in financial economics. It is widely recognized that high-frequency data are contaminated by microstructure noise (such as bid-ask spreads, late reporting, price discreteness, rounding errors or screen fighting), which is a challenge to the estimation and inference at the highest sampling frequencies. The *realized variance* (RV) - which is a sum of squared intraday returns - is biased and inconsistent when the high-frequency data are contaminated with noise. Recent work has therefore proposed a number of modifications of the RV that either reduce or, asymptotically, eliminate the impact of microstructure noise. Bandi & Russell (2005) derived the optimal sampling frequency of the RV, which minimizes its mean squared error. Their results show that the rule-of-thumb of using 5-minute returns to compute the RV tends to slightly understate the optimal sampling frequency for liquid equities. Zhang, Mykland & Aït-Sahalia (2005) proposed the subsampler, or two time-scales RV (TSRV), as the first consistent estimator of the IV in the presence of noise (for related work, see Kalnina & Linton (2006)). The TSRV converges at rate $N^{-1/6}$ and is a bias-corrected version of the RV, where the average of an increasing number of RV estimates across non-overlapping grids is used instead of a simple RV. Zhang (2005) used a multi-scale RV (MSRV), which has the efficient $N^{-1/4}$ rate of convergence. Barndorff-Nielsen, Hansen, Lunde & Shephard (2006*a*, 2006*b*) studied kernel-based estimators, where the noise is killed by incorporating realized autocovariances. Interestingly, the TSRV and MSRV are closely related to realized kernels. Large (2006) proposed an alternation estimator, which applies to asset markets where the price moves by a sequence of constant increments.

It has been suggested that the range is somewhat robust to common forms of microstructure noise (see, e.g., Alizadeh, Brandt & Diebold (2002)). Thus, range-based estimation of the IV is an interesting alternative in the presence of noise. Dijk & Martens (2006) studied the RRV with simulations and found it to be an accurate measure of the IV, which competes well against estimators that are robust to noise. However, if the RRV and RV are confined to the same sampling frequency, it was also reported that the RRV is the most biased statistic. Consequently, it is important to develop tools for bias-correcting the RRV. To our knowledge, no prior research has formally studied the impact of market microstructure noise on the RRV, and we fill that void here. We derive a theory for bias-correcting the RRV such that, under suitable conditions on the noise process, our new estimator is consistent for the IV with a mixed Gaussian central limit theorem (CLT).

The rest of the paper is organized as follows. In the next section, we state the semimartingale model and review quadratic variation. In section 3, we perturb the true price with microstructure noise and derive a robust realized range-based estimator of the IV. In section 4, we present some Monte Carlo simulations to inspect how accurate our estimator and distribution theory is for small sample sizes. In section 5, we present some empirical results based on high-frequency data of INTC and MSFT. A brief summary and some directions for future research conclude the paper in section 6.

2. A Brownian semimartingale

To fix ideas, we consider a continuous time log-price $p^* = (p_t^*)_{t \geq 0}$ that is defined on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$ and adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$. In an arbitrage-free frictionless market, the theory of financial economics implies that p^* must be of semimartingale form (see, e.g., Back (1991)). In this paper, we work with a Brownian semimartingale written as:

$$p_t^* = p_0^* + \int_0^t \mu_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u, \quad t \geq 0, \tag{2.1}$$

where $\mu = (\mu_t)_{t \geq 0}$ is locally bounded and predictable, $\sigma = (\sigma_t)_{t \geq 0}$ is a strictly positive process and $W = (W_t)_{t \geq 0}$ a standard Brownian motion. This process is also called a stochastic volatility model with drift (cf., e.g., Ghysels, Harvey & Renault (1996)).

To prove our CLTs, we will often work under some stronger assumptions on σ.

Assumption (V): *σ is everywhere invertible (V_1) and satisfies the equation:*

$$\sigma_t = \sigma_0 + \int_0^t \mu'_u \mathrm{d}u + \int_0^t \sigma'_u \mathrm{d}W_u + \int_0^t v'_u \mathrm{d}B'_u, \quad t \geq 0, \tag{$\mathbf{V_2}$}$$

where $\mu' = (\mu'_t)_{t \geq 0}$, $\sigma' = (\sigma'_t)_{t \geq 0}$ and $v' = (v'_t)_{t \geq 0}$ are càdlàg, with μ' also being locally bounded and *predictable,* $B' = (B'_t)_{t\geq 0}$ *is a Brownian motion, and* $W \perp\!\!\!\perp B'$ *(here* $A \perp\!\!\!\perp B$ *means that* A *and* B *are stochastically independent).*

This means that σ has its own Brownian semimartingale structure. Note the appearance of W in σ, which allows for leverage effects. V_2 is not necessary, but it simplifies the proofs considerably. A more general treatment, including the case where σ jumps, can be found in Barndorff-Nielsen, Graversen, Jacod, Podolskij & Shephard (2006). We rule out these technical details here, as they are not important to our exposition.

In what follows, we also make use of the concept of stable convergence in law.

Definition 1 *A sequence of random variables,* $(X_n)_{n\in\mathbb{N}}$, converges stably in law with limit X, *defined on an appropriate extension of* $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$, *if and only if for every* $\mathcal{F}$*-measurable, bounded random variable* Y *and any bounded, continuous function* g*, the convergence* $\lim_{n\to\infty} \mathbb{E}\left[Y g\left(X_n\right)\right] = \mathbb{E}\left[Y g\left(X\right)\right]$ *holds. We write* $X_n \overset{d_s}{\to} X$*, if* $(X_n)_{n\in\mathbb{N}}$ *converges stably in law to* X*.*

Stable convergence implies weak convergence, or convergence in law, which can be defined equivalently by taking $Y = 1$ (see, e.g., Rényi (1963) or Aldous & Eagleson (1978) for more details about the properties of stably converging sequences). The extension of this concept to stable convergence of processes is discussed in Jacod & Shiryaev (2003, pp. 512-518).

2.1. Quadratic variation

Crucial to the theory of semimartingales is the *quadratic variation* (QV). The QV is a key concept in high-frequency volatility and is fundamentally linked to financial risk. QV is defined as:

$$QV = \underset{n\to\infty}{\text{p-lim}} \sum_{i=1}^{n} \left(p^*_{t_i} - p^*_{t_{i-1}}\right)^2, \tag{2.2}$$

for any sequence of partitions $0 = t_0 < t_1 < \ldots < t_n = 1$ such that $\max_{1\leq i\leq n}\{t_i - t_{i-1}\} \to 0$ (see, e.g., Protter (2004)). In our setting, the QV is equal to the IV:

$$\int_0^1 \sigma_u^2 \mathrm{d}u. \tag{2.3}$$

3. Market microstructure noise

In practice, p^* is contaminated with microstructure noise, so there are deviations from the frictionless semimartingale framework (e.g., whereas changes in p^* are governed by continuous diffusive sample paths, the notion of a minimum tick size necessarily restricts changes in the observed price to discrete grids). We model this as

$$p_t = p_t^* + \eta_t, \tag{3.1}$$

where $p = (p_t)_{t \geq 0}$ denotes the observed price and $\eta = (\eta_t)_{t \geq 0}$ is i.i.d. noise with $\mathbb{E}(\eta_t) = 0$, $\mathbb{E}\left(\eta_t^2\right) = \omega^2$ and $\eta \perp\!\!\!\perp p^*$.

The i.i.d. assumption is not appropriate in continuous time (see, e.g., Kalnina & Linton (2006)), but Hansen & Lunde (2006) find little empirical evidence against it for liquid equities, when the sampling interval is above a minute. In our setting, however, the condition must hold down to the tick level, as the range is a functional of all the data within the sampling interval. Thus, we will view the i.i.d. assumption as an approximation here and attempt to relax it in future work. We note that the assumption has been dispensed with in some recent papers (cf., e.g., Aït-Sahalia, Mykland & Zhang (2006), Barndorff-Nielsen, Hansen, Lunde & Shephard (2006*a*), or Kalnina & Linton (2006)).

3.1. Realized variance

We assume that high-frequency data of p are recorded at the discrete points i/N for $i = 0, 1, \ldots, N$ with $N = mn$. The data partition the interval $[0, 1]$, which - for concreteness - is thought of as a trading day. Given these price data, we construct ultra high-frequency returns $r_{i\Delta',\Delta'} = p_{i/N} - p_{(i-1)/N}$, for $i = 1, \ldots, N$, where $\Delta' = 1/N$, and define the RV at sampling frequency N by setting

$$RV^N = \sum_{i=1}^{N} r_{i\Delta',\Delta'}^2. \tag{3.2}$$

Without microstructure noise, it follows that $RV^N \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u$ as $N \to \infty$, where we use "$\xrightarrow{p}$" to denote convergence in probability. Moreover, in the parametric setting RV^N is the ML estimator. The asymptotic distribution of the RV was derived in Jacod (1994), Jacod & Protter (1998), and Barndorff-Nielsen & Shephard (2002), and is given by

$$N^{1/2}\left(RV^N - \int_0^1 \sigma_u^2 \mathrm{d}u\right) \xrightarrow{d_s} MN\left(0, 2\int_0^1 \sigma_u^4 \mathrm{d}u\right), \tag{3.3}$$

where

$$\int_0^1 \sigma_u^4 \mathrm{d}u \tag{3.4}$$

is called the *integrated quarticity* (IQ). Note that the IQ, like the IV, is a latent variable being an integral of σ. Thus, the distribution theory in Equation (**3.3**) is infeasible, because it cannot be implemented. A feasible estimator of the IQ (in absence of noise) is the realized quarticity

$$RQ^N = \frac{N}{3} \sum_{i=1}^{N} r_{i\Delta',\Delta'}^4 \xrightarrow{p} \int_0^1 \sigma_u^4 \mathrm{d}u, \tag{3.5}$$

from which it follows that

$$\frac{N^{1/2}\left(RV^N - \int_0^1 \sigma_u^2 \mathrm{d}u\right)}{\sqrt{2RQ^N}} \xrightarrow{d} N(0,1), \tag{3.6}$$

where "$\xrightarrow{d}$" denotes convergence in law.

With i.i.d. noise RV^N has an bias of $2N\omega^2$ so $RV^N \xrightarrow{p} \infty$ as $N \to \infty$. This means that RV^N is not an appropriate estimator, although in general the sign and magnitude of the bias depend on the properties of η. The most simple solution to this problem is to avoid sampling at too high a frequency, that is to choose $n \ll N$, such that the resulting bias term $2n\omega^2$ can be ignored. This defines the sparsely sampled RV at sampling frequency n:

$$RV^n = \sum_{i=1}^{n} r_{i\Delta,\Delta}^2, \tag{3.7}$$

where $r_{i\Delta,\Delta} = p_{i/n} - p_{(i-1)/n}$ and $\Delta = 1/n$. In practice, the choice of n is often guided by volatility signatures, which is to calculate the time series average of RV^n for different n (cf., e.g., Fang (1996) or Andersen, Bollerslev, Diebold & Labys (2000)).

[INSERT TABLE 1 AND 2 ABOUT HERE]

In Table 1 and 2, we construct volatility signatures for some transaction data of Intel (INTC) and Microsoft (MSFT) that are further analyzed in our empirical section. The sample period is January 2, 2003 to December 31, 2004. The tables are based on daily estimates of RV^n. Note that here, as is standard in the literature, we use the sampling interval Δ (measured in minutes) so that n increases as time between observations decreases. If there was no microstructure noise, we would expect the average of RV^n to be independent of n. This is not what we see. INTC and MSFT trade almost down to the second, but it is clear that computing the RV every few seconds or so would lead to substantial bias.

3.2. Realized range-based variance

Christensen & Podolskij (2006*a*, 2006*b*) proposed the RRV (see, e.g., Parkinson (1980), Garman & Klass (1980), Rogers & Satchell (1991), or Dijk & Martens (2006) for related work on range-based volatility). The main idea of the RRV is to reduce the information loss of RV^n by replacing squared returns with squared ranges. Write

$$s_{p_{i\Delta},\Delta,m} = \max_{0 \leq s,t \leq m} \left(p_{\frac{i-1}{n}+\frac{t}{N}} - p_{\frac{i-1}{n}+\frac{s}{N}} \right), \tag{3.8}$$

for $i = 1, \ldots, n$ and set

$$RRV^{n,m} = \frac{1}{\lambda_{2,m}} \sum_{i=1}^{n} s^2_{p_{i\Delta},\Delta,m}, \tag{3.9}$$

where $\lambda_{r,m} = \mathbb{E}\left[\max_{0 \leq s,t \leq m} \left(W_{t/m} - W_{s/m}\right)^r\right]$. If $\eta = 0$, it holds that $RRV^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u$ as $n \to \infty$. To get a mixed normal distribution theory, we require Assumption **(V)** and the convergence $m \to c \in \mathbb{N} \cup \{\infty\}$ to obtain that

$$\sqrt{n}\left(RRV^{n,m} - \int_0^1 \sigma_u^2 \mathrm{d}u\right) \xrightarrow{d_s} MN\left(0, \Lambda_c \int_0^1 \sigma_u^4 \mathrm{d}u\right), \tag{3.10}$$

where $\Lambda_c = \lim_{m \to c} \Lambda_m$ and $\Lambda_m = \left(\lambda_{4,m} - \lambda_{2,m}^2\right)/\lambda_{2,m}^2$. Λ_m is decreasing in m and takes values between 2 ($m = 1$) and about 0.4 ($m \to \infty$). Thus, $RRV^{n,m}$ is more efficient than RV^n, whenever $m > 1$. However, if $\eta = 0$ nothing prevents us from constructing RV^N, which we know is asymptotically most efficient. Thus, it only makes sense to use $RRV^{n,m}$, when there are microstructure frictions.

3.3. Distributional assumption on the noise

In practice, $s_{p_{i\Delta},\Delta,m}$ is affected by $m + 1$ microstructure errors and the impact of the noise is severe. In Table 1 and 2 we confirm empirically the finding of Dijk & Martens (2006) that using a fixed n, $RRV^{n,m}$ is much more biased than RV^n.

This suggests that a bias-correction can improve upon $RRV^{n,m}$. It is not possible, however, to develop consistent, asymptotically mixed normal estimators of the IV, using the RRV, in the presence of a general i.i.d. microstructure noise. This is because the extreme value theory depends on the distribution of η. Thus, we need further assumptions on η. Our setup is formulated as Assumption (N).

Assumption (N): *η_t has density function*

$$P^{\eta_t} = \frac{1}{2}\left(\delta_\omega + \delta_{-\omega}\right), \tag{3.11}$$

where δ is the Dirac measure and ω is a positive constant.

This setup is very simple and we discuss various extensions of it to richer families of parametric densities below. We choose this setup due to its simplicity and because it works extremely well for the high-frequency data we investigate in our empirical application. Dijk & Martens (2006) have previously used this assumption in their simulation experiments, and we will loosely think of ω as a "half-spread".

3.4. Estimating the variance of the noise process

Now, we propose a robust RRV estimator of the IV in the presence of microstructure noise. The first step is to obtain a consistent estimate of ω. It turns out that, whereas RV^N is useless for estimating the IV, it can be useful for estimating the variance of the noise process, ω^2.

Lemma 1 *Suppose that p^* satisfies Equation* (**2.1**) *and that $p_t = p_t^* + \eta_t$, where η is i.i.d. with $\mathbb{E}(\eta_t) = 0$, $\mathbb{E}(\eta_t^2) = \omega^2$, and $\eta \perp\!\!\!\perp p^*$. Then it holds that*

$$\hat{\omega}_N^2 = \frac{RV^N}{2N} \xrightarrow{p} \omega^2, \tag{3.12}$$

and

$$N^{1/2}\left(\hat{\omega}_N^2 - \omega^2\right) \xrightarrow{d} N\left(0, \omega^4\right). \tag{3.13}$$

In the setting without drift ($\mu = 0$), the bias of $\hat{\omega}_N^2$ is $\int_0^1 \sigma_u^2 \mathrm{d}u/2N$, which can be large in practice in comparison to ω^2. Oomen (2005) suggested an alternative estimator based on the negative of the first-order sample autocovariance of returns.

Lemma 2 *Under the assumptions of Lemma 1, we have that*

$$\tilde{\omega}_N^2 = -\frac{1}{N-1}\sum_{i=1}^{N-1} r_{i\Delta',\Delta'} r_{(i+1)\Delta',\Delta'} \xrightarrow{p} \omega^2, \tag{3.14}$$

and

$$N^{1/2}\left(\tilde{\omega}_N^2 - \omega^2\right) \xrightarrow{d} N\left(0, 5\omega^4\right). \tag{3.15}$$

It is worth pointing out that $\hat{\omega}_N^2$ and $\tilde{\omega}_N^2$ are consistent estimators of ω^2 under general assumptions about η, and not only the distribution adopted here. In absence of drift, $\mathbb{E}\left(r_{i\Delta',\Delta'} r_{(i+1)\Delta',\Delta'}\right) = -\omega^2$ and all higher-order autocovariances are zero, leading to MA(1) dependence. $\tilde{\omega}_N^2$ is therefore unbiased, but its asymptotic variance is higher than that of $\hat{\omega}_N^2$ (this holds in general, see, e.g., Barndorff-Nielsen, Hansen, Lunde & Shephard (2006*a*)). In all instances, terms involving drift play a minor role, as the drift is $\mathrm{O}_p\left(N^{-1}\right)$. Here we base our analysis on $\hat{\omega}_N^2$, noting that the asymptotic distribution of our bias-corrected range-statistic is altered, if $\tilde{\omega}_N^2$ is used instead.

Using standard arguments, it holds that

$$\hat{\omega}_N \xrightarrow{p} \omega, \tag{3.16}$$

where $\hat{\omega}_N = \sqrt{\hat{\omega}_N^2}$.

3.5. Consistent estimation of the integrated variance

We then introduce the new realized range-based estimator of the IV:

$$RRV_{BC}^{n,m} = \frac{1}{\tilde{\lambda}_{2,m}} \sum_{i=1}^{n} \left(s_{p_{i\Delta},\Delta,m} - 2\hat{\omega}_N\right)^2, \tag{3.17}$$

where

$$\tilde{\lambda}_{r,m} = \mathbb{E}\left[\left| \max_{t:\eta_{\frac{t}{m}}=\omega,\ s:\eta_{\frac{s}{m}}=-\omega} \left(W_{\frac{t}{m}} - W_{\frac{s}{m}}\right)\right|^r\right], \tag{3.18}$$

with $1 \leq s, t \leq m$.

[INSERT FIGURE 1 ABOUT HERE]

In Figure 1, we plot simulated values of $\lambda_{2,m}$ and $\tilde{\lambda}_{2,m}$ for all m that integer divide 23,400. We do not know of any analytic formulas with which to express them. We used a counting variable in these simulations to keep track of the number of times, where the evaluation in Equation (**3.18**) resulted in the empty set on one of the indices. This is only relevant for very low m, for the probability of getting no positive or negative microstructure errors is 0.5^m. The figure is based on the average of all non-empty evaluations. This may call for not selecting m too low in practice to avoid unreliable normalizations. We use $m \geq 10$ in our simulations, which is sufficient to handle this. Note that $\tilde{\lambda}_{2,m}$ is independent of ω and $\lambda_{2,m} \to \lambda_2$ and $\tilde{\lambda}_{2,m} \to \lambda_2$ (where, in general, $\lambda_r = \lim_{m\to\infty} \lambda_{r,m} = \lim_{m\to\infty} \tilde{\lambda}_{r,m}$).

Theorem 1 *Suppose that the conditions of Lemma 1 and Assumption (N) hold. Then, as* $m, n \to \infty$

$$RRV_{BC}^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u. \tag{3.19}$$

Theorem 1 takes the form of a double asymptotics, in which both m and n are required to diverge to infinity. Intuitively, as $m \to \infty$ the observed minus true range (on small intervals) converges in probability to 2ω. Subtracting a consistent estimator of 2ω gives an asymptotically perfect bias-correction, and letting $n \to \infty$ we get the consistency for the IV, as in Christensen & Podolskij (2006*b*).

3.6. Asymptotic distribution of $RRV_{BC}^{n,m}$

Theorem 2 *Assume that the conditions of Theorem 1 are satisfied and that $m, n = \mathrm{O}\left(N^{1/2}\right)$. Set $n = c\sqrt{N}$ and $m = c^{-1}\sqrt{N}$. Moreover, we assume that there exists a Brownian motion $B'' = \left(B_t''\right)_{t\geq 0}$ defined on $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ with $p^* \perp\!\!\!\perp B''$ and such that*

$$\eta_{\frac{i}{N}} = \omega\left(2 \times 1_{\left\{\sqrt{N}\Delta_i^N B'' \geq 0\right\}} - 1\right), \qquad \textbf{(3.20)}$$

where $\Delta_i^N B'' = B_{i/N}'' - B_{(i-1)/N}''$. Then, the asymptotic distribution of $RRV_{BC}^{n,m}$ is given by:

$$N^{1/4}\left(RRV_{BC}^{n,m} - \int_0^1 \sigma_u^2 \mathrm{d}u\right) \xrightarrow{d_s} MN\left(0, \mathrm{avar}_{RRV_{BC}^{n,m}}\right), \qquad \textbf{(3.21)}$$

where

$$\mathrm{avar}_{RRV_{BC}^{n,m}} = \frac{\Lambda \int_0^1 \sigma_u^4 \mathrm{d}u}{c} + \frac{H_c^2 \omega^2}{4}, \qquad \textbf{(3.22)}$$

$\Lambda = \lim_{m\to\infty} \tilde{\Lambda}_m$,

$$\tilde{\Lambda}_m = \frac{\tilde{\lambda}_{4,m} - \tilde{\lambda}_{2,m}^2}{\tilde{\lambda}_{2,m}^2}, \qquad \textbf{(3.23)}$$

and

$$H_c = 4\sqrt{c}\frac{\lambda_1}{\lambda_2}\int_0^1 \sigma_u \mathrm{d}u, \qquad \textbf{(3.24)}$$

with

$$\int_0^1 \sigma_u \mathrm{d}u, \qquad \textbf{(3.25)}$$

being the integrated standard deviation (IS).

Remark 1 *The condition* **(3.20)** *ensures that both p^* and η are measurable with respect to the same type of filtration, which allows us to use existing CLTs for high-frequency data (see, e.g., Jacod & Shiryaev (2003)). This assumption has previously been used by Gloter & Jacod (2001*a*, 2001*b*).*

Remark 2 *$RRV_{BC}^{n,m}$ converges to the IV at rate $N^{-1/4}$, which is the fastest rate of convergence that can be obtained in this problem.*

Remark 3 *In* $\text{avar}_{RRV_{BC}^{n,m}}$*, there should be an additional (covariance) term*

$$\frac{\lambda_2 \kappa \omega H_c \int_0^1 \sigma_u^2 \mathrm{d}u}{\sqrt{c}}, \tag{3.26}$$

where $\kappa = \lim_{m\to\infty} \kappa_m$ *with*

$$\kappa_m = \mathbb{E}\left[\left|\max_{t:\tilde{\eta}_{\frac{t}{m}}=1, s:\tilde{\eta}_{\frac{s}{m}}=-1} \left(W_{\frac{t}{m}} - W_{\frac{s}{m}}\right)\right|^2 \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\frac{1}{2}\left(\tilde{\eta}_{\frac{i}{m}} - \tilde{\eta}_{\frac{i-1}{m}}\right)^2 - 1\right)\right], \tag{3.27}$$

where $1 \leq s, t \leq m$ *and* $\tilde{\eta}$ *has density* $\frac{1}{2}(\delta_1 + \delta_{-1})$*. However, by simple arguments it follows that* $\kappa = 0$*. Indeed, we have that*

$$\kappa_m = \mathbb{E}\left[\sup_{0\leq s,t\leq 1} (W_t - W_s)^2 \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \left(\frac{1}{2}\left(\tilde{\eta}_{\frac{i}{m}} - \tilde{\eta}_{\frac{i-1}{m}}\right)^2 - 1\right)\right] + \mathrm{o}_p(1) = \mathrm{o}_p(1). \tag{3.28}$$

Thus, this term drops out of the asymptotic variance.

While $\kappa_m \to 0$ as $m \to \infty$, it is not negligible for small m. We found the finite sample inference to improve a bit, when we included the additional term containing κ_m, as the distribution of $RRV_{BC}^{n,m}$ would otherwise be slightly overdispersed for the typical levels of m that we tend to select.

Remark 4 *The asymptotic conditional variance of* $RRV_{BC}^{n,m}$ *is minimized at*

$$c^* = \frac{\sqrt{\Lambda \int_0^1 \sigma_u^4 \mathrm{d}u}}{2\lambda_1 \lambda_2^{-1} \omega \int_0^1 \sigma_u \mathrm{d}u}, \tag{3.29}$$

with an optimal conditional variance equal to

$$\text{avar}^*_{RRV_{BC}^{n,m}} = 4\frac{\lambda_1}{\lambda_2}\sqrt{\Lambda}\omega \int_0^1 \sigma_u \mathrm{d}u \sqrt{\int_0^1 \sigma_u^4 \mathrm{d}u}. \tag{3.30}$$

If σ *is constant, this reduces to:*

$$\text{avar}^*_{RRV_{BC}^{n,m}} = 4\frac{\lambda_1}{\lambda_2}\sqrt{\Lambda}\omega\sigma^3 \approx 1.4972\omega\sigma^3. \tag{3.31}$$

Remark 5 *If we replace* $\hat{\omega}_N$ *with* $\tilde{\omega}_N$ *in the definition of* $RRV_{BC}^{n,m}$ *the conditional variance becomes*

$$\text{avar}_{RRV_{BC}^{n,m}} = \frac{\Lambda \int_0^1 \sigma_u^4 \mathrm{d}u}{c} + \frac{5H_c^2\omega^2}{4}. \tag{3.32}$$

Consequently, $\text{avar}_{RRV_{BC}^{n,m}}$ *is now minimized at*

$$c^* = \frac{\sqrt{\Lambda \int_0^1 \sigma_u^4 \mathrm{d}u}}{2\sqrt{5}\lambda_1 \lambda_2^{-1} \omega \int_0^1 \sigma_u \mathrm{d}u}, \tag{3.33}$$

and is equal to

$$\text{avar}^*_{RRV^{n,m}_{BC}} = 4\sqrt{5}\frac{\lambda_1}{\lambda_2}\sqrt{\Lambda}\omega \int_0^1 \sigma_u \mathrm{d}u \sqrt{\int_0^1 \sigma_u^4 \mathrm{d}u}. \tag{3.34}$$

When σ is constant we obtain

$$\text{avar}^*_{RRV^{n,m}_{BC}} = 4\sqrt{5}\frac{\lambda_1}{\lambda_2}\sqrt{\Lambda}\omega\sigma^3 \approx 3.3479\omega\sigma^3. \tag{3.35}$$

Note that the direct effect of a higher ω is to increase $\text{avar}_{RRV^{n,m}_{BC}}$ and lower the optimal sampling frequency n^*. However, in our setting this is partly compensated by an offsetting increase in m^*.

The conditional variance of $RRV^{n,m}_{BC}$ is infeasible, as it involves integrals of σ. To make the limit theory feasible, there are four quantities to estimate from the data. First, we construct noise robust estimates of the IQ and IS. Second, we must estimate ω and ω^2, which was discussed above.

It is quite simple to develop robust estimators of integrated power variation with our setup. We omit the general details here, but note that immediate corollaries are:

$$RRQ^{n,m}_{BC} = \frac{n}{\tilde{\lambda}_{4,m}} \sum_{i=1}^{n} \left(s_{p_{i\Delta,\Delta,m}} - 2\hat{\omega}_N\right)^4 \xrightarrow{p} \int_0^1 \sigma_u^4 \mathrm{d}u, \tag{3.36}$$

and

$$RRS^{n,m}_{BC} = \frac{1}{\tilde{\lambda}_{1,m}\sqrt{n}} \sum_{i=1}^{n} \left|s_{p_{i\Delta,\Delta,m}} - 2\hat{\omega}_N\right| \xrightarrow{p} \int_0^1 \sigma_u \mathrm{d}u, \tag{3.37}$$

as $m, n \to \infty$. The scaling of these estimators with n and $n^{-1/2}$ is required, since each raw term $(s_{p_{i\Delta,\Delta,m}} - 2\hat{\omega}_N)$ is $\mathrm{O}_p\left(n^{-1/2}\right)$. By using the properties of stable convergence in law (e.g., Jacod (1997)), we end up with a more standard convergence result:

$$\frac{N^{1/4}\left(RRV^{n,m}_{BC} - \int_0^1 \sigma_u^2 \mathrm{d}u\right)}{\sqrt{\widehat{\text{avar}}_{RRV^{n,m}_{BC}}}} \xrightarrow{d} N(0,1), \tag{3.38}$$

where $\widehat{\text{avar}}_{RRV^{n,m}_{BC}}$ is a consistent estimator of $\text{avar}_{RRV^{n,m}_{BC}}$. We can use this result to construct feasible confidence intervals for the IV in the presence of market microstructure noise.

3.7. Extensions of the basic framework

In this section, we will discuss various extensions of our framework.

3.7.1. The case with discrete noise

Our methodology also works for other discrete distributions with bounded support. However, we need at least minimal parametric assumptions on the distribution. For example, suppose that

$$P^{\eta_t} = \sum_{i=1}^{k} p_i \delta_{x_i}, \tag{3.39}$$

for some k and ordered points $x_1 < \ldots < x_k$ ($p_i > 0$ for all i). Since $\mathbb{E}(\eta_t) = 0$, we immediately get the two conditions $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} p_i x_i = 0$, which means that we require $2k-2$ further conditions to identify the parameters $p_1, \ldots, p_k, x_1, \ldots, x_k$. This can be done by using the method of moments, i.e. by computing the estimates $N^{-1} \sum_{i=1}^{N} r_{i\Delta',\Delta'}^{q}$, for $q = 2, \ldots, 2k-1$. Of course, if k is large we will encounter some problems in trying to solve this system of $2k$ non-linear equations. In addition, the moment estimators will be very small in practice for large q, which can be an empirical problem. However, once we have estimated the p's and x's, we can proceed as above by using instead $s_{p_{i\Delta,\Delta},m} - x_k + x_1$.

3.7.2. The case with continuous noise

To analyze the case where the noise is a continuous random variable, we assume that $\eta_t \sim U[-\nu, \nu]$, for some $\nu \in \mathbb{R}^+$. Here the microstructure noise has a uniform distribution on the interval $[-\nu, \nu]$. Now, we take $n = \mathrm{O}\left(N^{2/3-\delta}\right)$ and $m = \mathrm{O}\left(N^{1/3+\delta}\right)$, for some $\delta > 0$. Note that

$$\omega^2 = \frac{\nu^2}{3}, \tag{3.40}$$

with this model. Thus, we can define the estimator

$$\hat{\nu}_N = \sqrt{3\hat{\omega}_N} \xrightarrow{p} \nu, \tag{3.41}$$

as $N \to \infty$. We now study the bias-correction

$$RRV_{BC}^{n,m}(\nu) = \sum_{i=1}^{n} \left(s_{p_{i\Delta,\Delta},m} - 2\frac{m}{m+2}\hat{\nu}_N \right)^2. \tag{3.42}$$

The term $m/(m+2)$ is a small sample correction that disappears as $m \to \infty$.

Theorem 3 *The stochastic convergence*

$$RRV_{BC}^{n,m}(\nu) \xrightarrow{p} \frac{1}{3}\int_0^1 \sigma_u^2 \mathrm{d}u, \tag{3.43}$$

holds as $N \to \infty$.

Remark 6 *The convergence in probability can also be extended to (at least) any parametric η, where the density function has bounded support.*

We are not able to derive a CLT in this setting. The main idea of the proof of consistency is to replace the maximum of the increments of p with those of η plus the corresponding increments of p^*. The order of the error of this approximation is small for consistency, but blows up when we scale the statistic with $\sqrt{n}$ to prove the CLT. Further details are given in Lemma 5 in the Appendix.

3.7.3. The case with round-off errors

The lead example of microstructure noise is price discreteness, or round-off errors. Unfortunately, round-off errors is also the most difficult case to handle. In this section, we provide an idea of how to use realized range-based estimators in that situation.

The asymptotic theory for the RV in the presence of round-off errors was derived in Delattre & Jacod (1997). We follow their notation to call the accuracy of the measurements α_N, and note that Delattre & Jacod (1997) worked in the setting with $\alpha_N = \mathrm{O}\left(N^{-1/2}\right)$.[1] The observed rounded-off prices, $p^{(\alpha_N)}_{\frac{i-1}{N}}$, is then equal to $k\alpha_N$ for

$$k\alpha_N \leq p_{\frac{i-1}{N}} < (k+1)\,\alpha_N, \tag{3.44}$$

and $k \in \mathbb{Z}$. Alternatively

$$p^{(\alpha_N)}_{\frac{i-1}{N}} = p_{\frac{i-1}{N}} - \alpha_N \left\{ \frac{p_{\frac{i-1}{N}}}{\alpha_N} \right\}, \tag{3.45}$$

where $\{x\}$ is the fractional part of x.

Take $N = nm$ and assume that $\alpha_N = \mathrm{o}\left(n^{-1/2}\right)$. In practice, this means that once we are given an order of the accuracy, we choose m so large that $\alpha_N = \mathrm{o}\left(n^{-1/2}\right)$ holds.

We then define the realized range-statistic

$$RRV^{n,m}_{RO} = \frac{1}{\lambda_{2,m}} \sum_{i=1}^{n} s^2_{p_{i\Delta,\Delta,m,\alpha_N}}, \tag{3.46}$$

with

$$s_{p_{i\Delta,\Delta,m,\alpha_N}} = \max_{0 \leq s,t \leq m} \left(p^{(\alpha_N)}_{\frac{i-1}{n}+\frac{t}{N}} - p^{(\alpha_N)}_{\frac{i-1}{n}+\frac{s}{N}} \right). \tag{3.47}$$

[1] $\alpha_N = \mathrm{O}(N^{-1/2})$ is the only interesting case. If α_N is of a smaller order it does not influence the consistency, and if it is of a higher order it becomes very difficult to do anything (at least when $n = \mathrm{O}(N)$). See Delattre & Jacod (1997) for details.

Theorem 4 *It holds that*

$$RRV_{RO}^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u, \tag{3.48}$$

as $m, n \to \infty$.

The critical step with round-off errors is to choose n and m such that α_N is of a small order. Of course, an equivalent theorem can be stated for RV^n. It is not possible to use RV^N here, however, so because $RRV_{RO}^{n,m}$ exploits all the high-frequency data, we suspect that it works better.

3.7.4. Robust estimation of a jump component

It is of considerable interest in financial economics to know whether econometric models that have continuous sample paths, such as those governed by Equation (**2.1**), offer a satisfactory description of the data from a statistical viewpoint. There is a growing literature that shows how to distinguish between continuous sample path movements and jumps in asset prices (see, e.g., Barndorff-Nielsen & Shephard (2004*b*, 2006), Huang & Tauchen (2005), Jiang & Oomen (2005), or Christensen & Podolskij (2006*a*)). To see how our convergence results can be extended in this direction, we now assume that

$$p_t^* = p_0^* + \int_0^t \mu_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u + \sum_{i=1}^{N_t} J_i, \quad t \geq 0, \tag{3.49}$$

where, in addition to the components defined above, $N = (N_t)_{t \geq 0}$ is a finite-activity counting process and $J = (J_i)_{i=1,\ldots,N_t}$ represents the jumps in p^*. Then, it holds that

$$QV = \int_0^1 \sigma_u^2 \mathrm{d}u + \sum_{i=1}^{N_t} J_i^2. \tag{3.50}$$

It is well-known that the RV estimates the overall QV process and cannot be informative about the IV in these models. The same problem appears with the RRV (see, e.g., Christensen & Podolskij (2006*a*)). Thus, it is an even more ambitious goal to estimate the IV in the stochastic volatility, plus jump and noise models. We define the robust realized range-based bipower variation as

$$RBV_{BC}^{n,m} = \frac{1}{\tilde{\lambda}_{1,m}^2} \sum_{i=1}^{n-1} \left| s_{p_{i\Delta,\Delta,m}} - 2\hat{\omega}_N \right| \left| s_{p_{(i+1)\Delta,\Delta,m}} - 2\hat{\omega}_N \right|. \tag{3.51}$$

The no-noise version of $RBV_{BC}^{n,m}$ was studied in Christensen & Podolskij (2006*a*), where it was shown to be robust to finite-activity jumps.

We write $p_t = p_t^* + \eta_t$ again, where η is given by Assumption (N). Now, the convergences

$$RBV_{BC}^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u \qquad \text{and} \qquad RRV_{BC}^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u + \frac{1}{\lambda_2} \sum_{i=1}^{N_t} J_i^2, \tag{3.52}$$

hold as $m, n \to \infty$. Hence, $RBV_{BC}^{n,m}$ is a robust estimator of the IV in the presence of both stochastic volatility, jumps and noise. This type of result has recently been achieved in Podolskij & Vetter (2006). Here we also have that $\lambda_{2,m} \left(RRV_{BC}^{n,m} - RBV_{BC}^{n,m} \right) \xrightarrow{p} \sum_{i=1}^{N_t} J_i^2$ as $m, n \to \infty$, and properly armed with an asymptotic distribution theory under the null hypothesis of no jumps, we should in principle be able to use this result to draw noise robust inference about the jump process. This topic is left for future research.

4. Simulation study

In this section, we look at the bias-correction with simulated data to evaluate the finite sample accuracy of $RRV_{BC}^{n,m}$. Moreover, we inspect how well the first-order approximation offered by Equation (**3.38**) works for the distributions that arise in sample sizes of practical relevance. We simulate the model:

$$\begin{aligned} \mathrm{d}p_t^* &= \sigma_t \mathrm{d}W_t, \\ \mathrm{d}\ln \sigma_t^2 &= \theta(\xi - \ln \sigma_t^2)\mathrm{d}t + \gamma \mathrm{d}B_t, \end{aligned} \tag{4.1}$$

where W and B are Brownian motions, $W \perp\!\!\!\perp B$, and (θ, ξ, γ) is a parameter vector. Here the log-variance is a mean reverting Ornstein-Uhlenbeck process with mean ξ, mean reversion θ and volatility γ.

4.1. Simulation design

We create 100,000 repetitions of the bivariate system in Equation (**4.1**) using an Euler approximation and $N = 100, 200, 300, 450, 600, 900, 1200, 1500$. The parameters $(\theta, \xi, \gamma) = (0.032, -0.631, 0.115)$ are taken from Andersen, Benzoni & Lund (2002). The initial conditions are set at $p_0^* = 0$ and $\ln \sigma_0^2 = \xi$. We cloak p^* with i.i.d. noise, using Assumption (N) and $\omega^2 = 0.002$. This is a reasonable level for liquid equities (see, e.g., the web appendix of Barndorff-Nielsen, Hansen, Lunde & Shephard (2006*a*)).

To estimate ω^2, we use $\tilde{\omega}_N^2$ from Lemma 2. In our initial design, we used $\hat{\omega}_N^2$ from Lemma 1, but this estimator has a severe upward bias for most of the sample sizes considered here. We therefore confine our analysis to $\tilde{\omega}_N^2$, except when it is negative where we switch to $\hat{\omega}_N^2$. The optimal partition of the high-frequency data is applied under the additional constraint that $m \geq 10$. We construct $RRV_{BC}^{n,m}$, $RRQ_{BC}^{n,m}$, and $RRS_{BC}^{n,m}$ estimates based on the resulting partition. Inspections of the simulations show that when N is small, the bound on m is almost always hit. Thus, our results might be seen as conservative to the extent that we are using an estimator with an inoptimal variance.

4.2. Results

[INSERT TABLE 3 ABOUT HERE]

In Table 3, we report the finite sample distributions of three asymptotic pivots. Panel A is for the standardized $RRV_{BC}^{n,m}$ in the infeasible setting, where the $\text{avar}_{RRV_{BC}^{n,m}}$ is known a priori. In Panel B, we estimate the $\text{avar}_{RRV_{BC}^{n,m}}$ from the simulated data, and Panel C reports on the feasible log-based inference, using the delta method to conclude that

$$N^{1/4}\left(\ln RRV_{BC}^{n,m} - \ln \int_0^1 \sigma_u^2 \text{d}u\right) \xrightarrow{d} MN\left(0, \frac{\text{avar}_{RRV_{BC}^{n,m}}}{\left(\int_0^1 \sigma_u^2 \text{d}u\right)^2}\right). \qquad \textbf{(4.2)}$$

Panel A shows that, although the convergence settles, it takes some time for the asymptotics to kick in. There is a substantial distortion for small N. In Panel B, we see that replacing the $\text{avar}_{RRV_{BC}^{n,m}}$ with a consistent estimator makes the approximation a bit worse and changes the skewness. The log-transformation in Panel C improves upon the raw distribution theory of $RRV_{BC}^{n,m}$ and is a good description for the sampling variation of $\ln RRV_{BC}^{n,m}$ already at $N = 1,200 - 1,500$. It is preferable to use the log-based distribution theory in practice to construct confidence intervals for the IV. This also has the virtue of imposing non-negativity on the confidence bands.

[INSERT FIGURE 2 ABOUT HERE]

To compare $RRV_{BC}^{n,m}$ to some alternative estimators, Figure 2 reports on the root mean squared error (RMSE) of $RRV_{BC}^{n,m}$, $RRV^{n,m}$, RV^n, RV^{n^*}, and the TSRV. Here RV^{n^*} is the RV computed at an optimal sampling frequency (see, e.g., Bandi & Russell (2005)). $RRV^{n,m}$ and RV^n are based on the partition of $RRV_{BC}^{n,m}$. We calculate the TSRV by using regular allocation of the data and an optimal number of subgrids (K^*) found by use of the automatic selection formula in Zhang et al. (2005), which takes

$$K^* = \left(12 \frac{\omega^4}{\int_0^1 \sigma_u^4 \text{d}u}\right)^{1/3} N^{2/3}. \qquad \textbf{(4.3)}$$

We also apply a finite sample correction to the TSRV and denote the resulting statistic by $TSRV\,(K,J)-aa$ (to conform with our empirical work).

The plot shows ln(RMSE) to ease the interpretation. Looking at the figure, we note that $RRV^{n,m}$ is severely affected by microstructure noise and much more than RV^n, which is in line with our developments from above. The RMSE of $RRV^{n,m}$ is higher than the RMSE of $RRV_{BC}^{n,m}$ by the time $N = 200$, so the bias-correction is paramount. An appealing feature is that $RRV_{BC}^{n,m}$ has the lowest RMSE, and hence is also more efficient than the subsampler, for $N \geq 300$. The RMSE of $RRV_{BC}^{n,m}$ and $TSRV\,(K,J)-aa$ continues to decline in N, reflecting their consistency, but the relative RMSE of $TSRV\,(K,J)-aa$ to $RRV_{BC}^{n,m}$ increases with N, due to the inefficient $N^{-1/6}$ convergence rate of the TSRV.

5. Bias-correction using empirical data

The bias-corrected range-statistic is applied to some high-frequency data for Intel (INTC) and Microsoft (MSFT). We extract transaction data from the TAQ database for the sample period January 2, 2003 - December 31, 2004. The raw data were preprocessed and screened for outliers using standard filtering algorithms. We compute the $RRV_{BC}^{n,m}$, $RRV^{n,m}$, RV^n and TSRV of Aït-Sahalia et al. (2006), the latter being robust to serial dependence in η. We follow the recommendations in Barndorff-Nielsen, Hansen, Lunde & Shephard (2006*a*) to implement the TSRV, subsampling 5-minute returns and doing the area-adjustment to correct a slight downward biased. In addition, to avoid dependencies in η that show up at higher frequencies (e.g., Hansen & Lunde (2006)), we estimate ω^2 on a day-by-day basis using the average $\hat{\omega}_N^2$ from 60 subsampled grids of 1-minute returns.

To implement $RRV_{BC}^{n,m}$, we first calculated a volatility signature of $RRV_{BC}^{n,m}$ and found that series to be essentially flat down to sampling intervals of about 1 - 2 minutes. Then $RRV_{BC}^{n,m}$ starts increasing a bit, as we move below this interval length. This can be due to a number of things, including a misspecified noise process that can have a bigger impact at higher sampling frequencies or, as already discussed, problems associated with too small m. In principle, both issues may contribute to rendering our bias-correction inadequate at the highest frequencies.

Overall, our preliminary analysis suggests that our proposed bias-correction works extremely well in terms of controlling the impact of noise on the range for sampling intervals above 1 - 2 minutes. In general, however, for the levels of ω that we observe in the data, the optimal sampling frequency of $RRV_{BC}^{n,m}$ would imply time intervals of about 30 - 45 seconds. In view of this, we therefore decided not to use the optimal sampling frequency for these equities but instead implement $RRV_{BC}^{n,m}$ at the 2-minute interval, which is the highest we feel comfortable using.

In Table 1 and 2, we show some summary statistics for the IV estimators of INTC and MSFT. We see that $RRV_{BC}^{n,m}(2\text{mn})$ is around the level of the RV computed at the 5 - 30 minute frequency, which is what we would expect if the bias-correction works, as the RV should not suffer from a severe upward bias at these frequencies, in sharp contrast to the behavior of $RV(1 \text{ tick})$. It is worth noting that $RRV_{BC}^{n,m}$ has a weaker serial dependence than $RRV^{n,m}$, but that it is still quite persistent. Moreover, the (time series) standard deviation of $RRV_{BC}^{n,m}$ is much smaller than those of the alternative estimators reported here, pointing towards a greater sampling stability.

[INSERT FIGURE 3 AND 4 ABOUT HERE]

To illustrate the use of our distribution theory, we extract data for July, 2004 and plot 95% confidence intervals for $\sqrt{\int_0^1 \sigma_u^2 \mathrm{d}u}$ of INTC and MSFT in Figure 3 and 4. Here we focus on a standard deviation-type of volatility. The limit theory is based on a second application of the delta method, using that

$$N^{1/4}\left(\sqrt{RRV_{BC}^{n,m}} - \sqrt{\int_0^1 \sigma_u^2 \mathrm{d}u}\right) \xrightarrow{d} MN\left(0, \frac{\text{avar}_{RRV_{BC}^{n,m}}}{4\int_0^1 \sigma_u^2 \mathrm{d}u}\right), \tag{5.1}$$

and likewise for $\sqrt{RV^n}$. In this plot, RV^n is based on 5-minute sampling. We note that the point estimates of the $\sqrt{RV^n}$, $\sqrt{TSRV\,(K,J) - aa}$ and $\sqrt{RRV_{BC}^{n,m}}$ series are often close, whereas $\sqrt{RRV^{n,m}}$ is significantly higher and often lies outside the 95% confidence regions. The confidence intervals tend to overlap and those based on $\sqrt{RRV_{BC}^{n,m}}$ are most of the times substantially smaller than those of $\sqrt{RV^n}$. A minor caveat here is that it is very difficult to estimate the IQ in the presence of noise. As noted in Barndorff-Nielsen, Hansen, Lunde & Shephard (2006*a*) there is no research which has solved this problem.[2] We used a conservative sampling interval of 30 minutes for both IQ estimators in these plots to avoid the worst bite of microstructure noise. This, of course, increases their sampling errors and can lead to larger swings in the day-to-day confidence intervals, although we do get some compensation here by focusing on estimation of $\sqrt{\int_0^1 \sigma_u^2 \mathrm{d}u}$ due to the natural scaling with the IV in the asymptotic conditional variance.

[2] Podolskij & Vetter (2006) is a recent paper that has made some progress on this problem.

6. Conclusions and directions for future work

In this paper, we proposed a realized range-based estimator of the IV that is robust to simple forms of microstructure noise. We derived a bias-correction to the range-statistic, such that the new estimator is consistent and asymptotically mixed Gaussian. Moreover, we showed how to optimally divide the high-frequency data to minimize its asymptotic conditional variance.

The paper highlights the potential that range-based estimation of the IV can exhibit under suitable conditions on the noise. On the one hand, we had to impose some parametric assumptions on the noise process to develop our bias-correction. On the other hand, we feel that our empirical results show that the proposed bias-correction does a good job for the transaction data analyzed here, provided we do not base our estimation and inference on the highest sampling frequencies.

In future work, we intend to look at realized range-based estimation of the *integrated covariation*, which is a key concept in financial economics. The interested reader is referred to, e.g., Barndorff-Nielsen & Shephard (2004*a*), Hayashi & Yoshida (2005), Brandt & Diebold (2006), Griffin & Oomen (2006) or Sheppard (2006) for some recent work in this exciting area.

A. **Appendix**

We assume, without loss of generality, that μ, σ, μ', σ', and v' are bounded (see, e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij & Shephard (2006)). In the following, we also use the notation

$$I_i = \left\{ (t,s) \mid t = \frac{i-1}{n} + \frac{k}{N} \text{ with } \eta_t = \omega,\ s = \frac{i-1}{n} + \frac{j}{N} \text{ with } \eta_s = -\omega, 1 \leq k, j \leq m \right\}.$$

The next Lemma provides a representation of $RRV_{BC}^{n,m}$.

Lemma 3 *It holds that*

$$RRV_{BC}^{n,m} = \frac{1}{\tilde{\lambda}_{2,m}} \sum_{i=1}^{n} \left(\max_{(t,s)\in I_i} \left(p^*_{\frac{i-1}{n}+\frac{t}{N}} - p^*_{\frac{i-1}{n}+\frac{s}{N}} \right) + 2\left(\hat{\omega}_N - \omega\right) \right)^2 + o_p\left(n^{-1/2}\right). \tag{A.1}$$

Proof

Write

$$s_{p_{i\Delta},\Delta,m} = \max\left\{ s^{(1)}_{p_{i\Delta},\Delta,m}, s^{(2)}_{p_{i\Delta},\Delta,m}, s^{(3)}_{p_{i\Delta},\Delta,m}, s^{(4)}_{p_{i\Delta},\Delta,m} \right\},$$

where

$$s^{(1)}_{p_{i\Delta},\Delta,m} = \max_{t:\eta_{\frac{i-1}{n}+\frac{t}{N}}=\omega,\ s:\eta_{\frac{i-1}{n}+\frac{s}{N}}=-\omega} \left(p^*_{\frac{i-1}{n}+\frac{t}{N}} - p^*_{\frac{i-1}{n}+\frac{s}{N}} \right) + 2\omega,$$

$$s^{(2)}_{p_{i\Delta},\Delta,m} = \max_{t:\eta_{\frac{i-1}{n}+\frac{t}{N}}=-\omega,\ s:\eta_{\frac{i-1}{n}+\frac{s}{N}}=\omega} \left(p^*_{\frac{i-1}{n}+\frac{t}{N}} - p^*_{\frac{i-1}{n}+\frac{s}{N}} \right) - 2\omega,$$

$$s^{(3)}_{p_{i\Delta},\Delta,m} = \max_{t:\eta_{\frac{i-1}{n}+\frac{t}{N}}=\omega,\ s:\eta_{\frac{i-1}{n}+\frac{s}{N}}=\omega} \left(p^*_{\frac{i-1}{n}+\frac{t}{N}} - p^*_{\frac{i-1}{n}+\frac{s}{N}} \right),$$

$$s^{(4)}_{p_{i\Delta},\Delta,m} = \max_{t:\eta_{\frac{i-1}{n}+\frac{t}{N}}=-\omega,\ s:\eta_{\frac{i-1}{n}+\frac{s}{N}}=-\omega} \left(p^*_{\frac{i-1}{n}+\frac{t}{N}} - p^*_{\frac{i-1}{n}+\frac{s}{N}} \right),$$

for $1 \leq s, t \leq m$. It suffices to show that

$$P\left(s^{(1)}_{p_{i\Delta},\Delta,m} \leq s^{(k)}_{p_{i\Delta},\Delta,m} \right) = o\left(n^{-3/2}\right),$$

for $k = 2, 3, 4$. We prove this result for $k = 3$ (the rest can be shown analogously). For all $p > 0$, Burkholder's inequality yields:

$$P\left(s^{(1)}_{p_{i\Delta},\Delta,m} \leq s^{(3)}_{p_{i\Delta},\Delta,m} \right) \leq P\left(s_{p^*_{i\Delta},\Delta} \geq 2\omega \right) \leq C\frac{n^{-p/2}}{\omega^{p/2}},$$

This completes the proof. ■

Next, notice that in view of Lemma 3 we get the decomposition

$$RRV_{BC}^{n,m} = V_1^n + V_2^n + V_3^n + \mathrm{o}_p\Big(n^{-1/2}\Big),$$

with

$$V_1^n = \frac{1}{\tilde{\lambda}_{2,m}} \sum_{i=1}^{n} \left| \max_{(t,s)\in I_i} (p_t^* - p_s^*) \right|^2,$$

$$V_2^n = \frac{4}{\tilde{\lambda}_{2,m}} (\hat{\omega}_N - \omega) \sum_{i=1}^{n} \max_{(t,s)\in I_i} (p_t^* - p_s^*),$$

$$V_3^n = \frac{4n}{\tilde{\lambda}_{2,m}} (\hat{\omega}_N - \omega)^2.$$

Using $(\hat{\omega}_N - \omega) = \mathrm{O}_p\left(N^{-1/2}\right)$, it follows that

$$V_1^n = \mathrm{O}_p(1),$$

$$V_2^n = \mathrm{O}_p\Big(m^{-1/2}\Big),$$

$$V_3^n = \mathrm{O}_p\left(N^{-1}\right). \tag{A.2}$$

This means V_3^n is negligible for the consistency and the CLT, whereas V_2^n is negligible for consistency only, but it appears in the CLT (recall that $n, m = \mathrm{O}\left(N^{1/2}\right)$).

A.1. **Proof of Theorem 1**

With these preliminary steps, the decomposition

$$RRV_{BC}^{n,m} = V_1^n + \mathrm{o}_p(1),$$

holds. Hence, the convergence

$$RRV_{BC}^{n,m} \xrightarrow{p} \int_0^1 \sigma_u^2 \mathrm{d}u,$$

is shown as in Christensen & Podolskij (2006*b*). ■

A.2. **Proof of Theorem 2**

In order for us to prove the CLT, we need the following result, which will be shown later.

Theorem 5 *If Assumption (V) and* **(3.20)** *are satisfied, then we have*

$$\begin{pmatrix} \sqrt{n}\left(V_1^n - \int_0^1 \sigma_u^2 \mathrm{d}u\right) \\ \sqrt{N}\left(\hat{\omega}_N^2 - \omega\right) \end{pmatrix} \xrightarrow{d_s} \int_0^1 \Sigma_s^{1/2} \mathrm{d}B'_s, \tag{A.3}$$

where B' is a 2-dimensional Brownian motion defined on an extension of the filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ and is independent of the σ-field $\mathcal{F}$. The matrix Σ is defined by

$$\Sigma_s = \begin{pmatrix} \Lambda\sigma_s^4 & 0 \\ 0 & \omega^4 \end{pmatrix}.$$

With Theorem 5 at hand, we able to prove the CLT. First, observe that the estimations in (A.2) imply the decomposition

$$N^{1/4}\left(RRV_{BC}^{n,m} - \int_0^1 \sigma_u^2 \mathrm{d}u\right) = V_{n,m}(1) + V_{n,m}(2) + \mathrm{o}_p(1),$$

where

$$V_{n,m}(1) = \frac{\sqrt{n}}{\sqrt{c}}\left(V_1^n - \int_0^1 \sigma_u^2 \mathrm{d}u\right),$$

$$V_{n,m}(2) = \sqrt{cm}V_2^n.$$

The second term admits the stochastic expansion

$$V_{n,m}(2) = \sqrt{N}H_{c,m}\left(\hat{\omega}_N - \omega\right) + \mathrm{o}_p(1),$$

where

$$H_{c,m} = 4\sqrt{c}\frac{\tilde{\lambda}_{1,m}}{\tilde{\lambda}_{2,m}}\int_0^1 \sigma_u \mathrm{d}u.$$

Now the CLT follows from Theorem 5 by an application of the delta method for the function $g(x,y) = \frac{x}{\sqrt{c}} + H_{c,m}\sqrt{y}$. ■

Proof of Theorem 5

We prove Theorem 5 in several steps. First, we show the next Lemma.

Lemma 4 *Assume that conditions (V) and* **(3.20)** *are satisfied. Set*

$$U_{n,m} = \sqrt{n}\sum_{i=1}^{n} \sigma_{\frac{i-1}{n}}^2 \left(\frac{1}{\tilde{\lambda}_{2,m}}\left|\max_{(t,s)\in I_i}(W_t - W_s)\right|^2 - \frac{1}{n}\right).$$

Then we have that

$$\begin{pmatrix} U_{n,m} \\ \sqrt{N}\left(\omega^2 - \hat{\omega}_N^2\right) \end{pmatrix} \xrightarrow{d_s} \int_0^1 \Sigma_s^{1/2} \mathrm{d}B'_s. \tag{A.4}$$

Proof

We define the quantities

$$\xi_i^{n,m} = \frac{1}{\sqrt{N}} \sum_{k=1}^{m} \left(\frac{1}{2} \left(\eta_{\frac{i-1}{n}+\frac{k}{N}} - \eta_{\frac{i-1}{n}+\frac{k-1}{N}} \right)^2 - \omega^2 \right),$$

$$\zeta_i^{n,m} = \sqrt{n}\sigma^2_{\frac{i-1}{n}} \left(\frac{1}{\tilde{\lambda}_{2,m}} \Big| \max_{(t,s)\in I_i} (W_t - W_s) \Big|^2 - \frac{1}{n} \right),$$

to obtain

$$U_{n,m} = \sum_{i=1}^{n} \zeta_i^{n,m} \qquad \text{and} \qquad \sqrt{N}\left(\hat{\omega}_N^2 - \omega^2\right) = \sum_{i=1}^{n} \xi_i^{n,m} + \mathrm{o}_p(1).$$

As the representation (**3.20**) holds, Theorem IX 7.28 in Jacod & Shiryaev (2003) is applicable for the vector $(\sum_{i=1}^n \zeta_i^{n,m}, \sum_{i=1}^n \xi_i^{n,m})^T$ (here T means transpose). Note the identities

$$\mathbb{E}\left[\zeta_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0, \qquad \mathbb{E}\left[(\zeta_i^{n,m})^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \tilde{\Lambda}_m \frac{1}{n} \sigma^4_{\frac{i-1}{n}},$$

and

$$\mathbb{E}\left[\xi_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0, \qquad \mathbb{E}\left[(\xi_i^{n,m})^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \frac{1}{n}\omega^4.$$

It follows that

$$\sum_{i=1}^{n} \mathbb{E}\left[(\zeta_i^{n,m})^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} \Lambda \int_0^1 \sigma_u^4 \mathrm{d}u, \qquad \sum_{i=1}^{n} \mathbb{E}\left[(\xi_i^{n,m})^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] = \omega^4.$$

Note that since $W \perp\!\!\!\perp B$ and $m \to \infty$, we get

$$\begin{aligned} \sum_{i=1}^{n} \mathbb{E}\left[\zeta_i^{n,m}\xi_i^{n,m} \mid \mathcal{F}_{\frac{i-1}{n}}\right] &= \sqrt{n} \sum_{i=1}^{n} \frac{1}{\tilde{\lambda}_{2,m}} \sigma^2_{\frac{i-1}{n}} \mathbb{E}\left[\xi_i^{n,m} \left| \max_{(s,t)\in I_i} (W_t - W_s)\right|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \\ &= \sqrt{n} \sum_{i=1}^{n} \frac{1}{\tilde{\lambda}_{2,m}} \sigma^2_{\frac{i-1}{n}} \mathbb{E}\left[\xi_i^{n,m} \sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]} (W_t - W_s)^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] + \mathrm{o}_p(1) \\ &= \mathrm{o}_p(1). \end{aligned}$$

Next, let $Z = W$ or $Z = B$. Since $(W,B) \overset{d}{=} -(W,B)$, we get

$$\mathbb{E}\left[\zeta_i^{n,m}\Delta_i^n Z \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0, \qquad \mathbb{E}\left[\xi_i^{n,m}\Delta_i^n Z \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0.$$

Finally, let $N = (N_t)_{t\in[0,1]}$ be a bounded martingale on $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,1]}, \mathcal{P}\right)$, which is orthogonal to W and B (i.e., with quadratic covariation $[W,N]_t = [B,N]_t = 0$, almost surely). By standard arguments (see, e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij & Shephard (2006)), we obtain the identity

$$\mathbb{E}\left[\zeta_i^{n,m}\Delta_i^n N \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0, \qquad \mathbb{E}\left[\xi_i^{n,m}\Delta_i^n N \mid \mathcal{F}_{\frac{i-1}{n}}\right] = 0.$$

Now, the stable convergence in law follows by Theorem IX 7.28 in Jacod & Shiryaev (2003). ■

We proceed with the proof of Theorem 2. Using the arguments of Christensen & Podolskij (2006*b*), Theorem 2 can be deduced from Lemma 4 and the condition

$$\sqrt{n}\sum_{i=1}^{n}\mathbb{E}\left[\left|\max_{(t,s)\in I_i}\left(p_t^* - p_s^*\right)\right|^2 - \sigma_{\frac{i-1}{n}}^2\left|\max_{(t,s)\in I_i}\left(W_t - W_s\right)\right|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \xrightarrow{p} 0. \tag{A.5}$$

From Christensen & Podolskij (2006*b*), we obtain the approximation

$$\begin{aligned}
&\sqrt{n}\sum_{i=1}^{n}\mathbb{E}\left[\left|\max_{(t,s)\in I_i}\left(p_t^* - p_s^*\right)\right|^2 - \sigma_{\frac{i-1}{n}}^2\left|\max_{(t,s)\in I_i}\left(W_t - W_s\right)\right|^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] \\
&= 2\sqrt{n}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}\max_{(t,s)\in I_i}\left(W_t - W_s\right)\left(\max_{(t,s)\in I_i}\left(p_t^* - p_s^*\right) - \sigma_{\frac{i-1}{n}}\max_{(t,s)\in I_i}\left(W_t - W_s\right)\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] + o_p(1) \\
&= 2\sqrt{n}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(W_t - W_s\right)\left(\max_{(s,t)\in I_i}\left(p_t^* - p_s^*\right) - \sigma_{\frac{i-1}{n}}\max_{(t,s)\in I_i}\left(W_t - W_s\right)\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] + o_p(1) \\
&= W_{n,m} + o_p(1),
\end{aligned}$$

where the second equality follows because $m \to \infty$. Next, we define the pair

$$\left(t_i^*(W,B), s_i^*(W,B)\right) = \underset{(t,s)\in I_i}{\arg\sup}\sqrt{n}\left(W_t - W_s\right), \tag{A.6}$$

as a functional of (W,B). It is simple to deduce that

$$\left(t_i^*(-W,-B), s_i^*(-W,-B)\right) = \left(s_i^*(W,B), t_i^*(W,B)\right). \tag{A.7}$$

Following Christensen & Podolskij (2006*b*), we find that

$$W_{n,m} = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}\left[\sigma_{\frac{i-1}{n}}\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(W_t - W_s\right)g_i\left(t_i^*(W,B), s_i^*(W,B)\right) \mid \mathcal{F}_{\frac{i-1}{n}}\right] + o_p(1),$$

where the function g_i is given by

$$g_i(s,t) = n\int_s^t \mu'_{\frac{i-1}{n}}\mathrm{d}u + n\int_s^t\left\{\sigma'_{\frac{i-1}{n}}\left(W_u - W_{\frac{i-1}{n}}\right) + v'_{\frac{i-1}{n}}\left(V_u - V_{\frac{i-1}{n}}\right)\right\}\mathrm{d}W_u.$$

As a consequence of (A.7), $g_i\left(t_i^*\left(W,B\right),s_i^*\left(W,B\right)\right)$ is an odd functional of (W,B,V). Moreover, $(W,B,V)\overset{d}{=}-\left(W,B,V\right)$, which means that

$$\mathbb{E}\left[\sigma_{\frac{i-1}{n}}\sup_{s,t\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(W_t-W_s\right)g_i\left(t_i^*\left(W,B\right),s_i^*\left(W,B\right)\right)\mid\mathcal{F}_{\frac{i-1}{n}}\right]=0,$$

and the proof is complete. ∎

A.3. **Proof of Theorem 3**

The next Lemma helps to separate the influence of η and p^* on $s_{p_{i\Delta},\Delta,m}$. A deterministic version of the Lemma (including a proof) can be found as Lemma 10 in Christensen & Podolskij (2006*a*).

Lemma 5 *Let s^*/N and t^*/N denote the almost surely unique points in the interval $\left[\frac{i-1}{n},\frac{i}{n}\right]$, where minimum and maximum of the process η are attained. Then*

$$n^{1/2}\left(s_{p_{i\Delta},\Delta,m}-\max_{\frac{s}{N},\frac{t}{N}\in\left[\frac{i-1}{n},\frac{i}{n}\right]}\left(\eta_{\frac{t}{N}}-\eta_{\frac{s}{N}}\right)\right)-n^{1/2}\left(p^*_{\frac{t^*}{N}}-p^*_{\frac{s^*}{N}}\right)=h_{in},$$

holds with $\mathbb{E}\left[\left|h_{in}\right|^q\right]=\mathrm{o}\left(1\right)$ for any $q>0$, uniformly in i.

Using Lemma 5, we conclude that

$$RRV_{BC}^{n,m}\left(\nu\right)=Z_1^n+Z_2^n+Z_3^n+\mathrm{o}_p\left(1\right),$$

with

$$Z_1^n=\sum_{i=1}^n\left(\left(\eta_{\frac{t^*}{N}}-\eta_{\frac{s^*}{N}}\right)-2\frac{m}{m+2}\hat{\nu}_N\right)^2,$$

$$Z_2^n=2\sum_{i=1}^n\left(\left(\eta_{\frac{t^*}{N}}-\eta_{\frac{s^*}{N}}\right)-2\frac{m}{m+2}\hat{\nu}_N\right)\left(p^*_{\frac{t^*}{N}}-p^*_{\frac{s^*}{N}}\right),$$

$$Z_3^n=\sum_{i=1}^n\left(p^*_{\frac{t^*}{N}}-p^*_{\frac{s^*}{N}}\right)^2.$$

Now, simple calculations show that

$$\left(\eta_{\frac{t^*}{N}}-\eta_{\frac{s^*}{N}}\right)-2\frac{m}{m+2}\nu=\mathrm{O}_p\left(m^{-1}\right).$$

Moreover, since $\hat{\nu}_N-\nu=\mathrm{O}_p\left(N^{-1/2}\right)$, a usage of Burkholder's inequality yields

$$Z_1^n=\mathrm{O}_p\left(\frac{n}{m^2}\right)=\mathrm{o}_p\left(1\right),$$

$$Z_2^n = \mathrm{O}_p\left(\frac{\sqrt{n}}{m}\right) = \mathrm{o}_p(1),$$

$$Z_3^n = \mathrm{O}_p(1).$$

To show the stochastic convergence

$$Z_3^n \xrightarrow{p} \frac{1}{3}\int_0^1 \sigma_u^2 \mathrm{d}u,$$

we use the arguments in, e.g., Christensen & Podolskij (2006*b*) to deduce that

$$Z_3^n = \sum_{i=1}^n \sigma_{\frac{i-1}{n}}^2 \mathbb{E}\left[\left(W_{\frac{t^*}{N}} - W_{\frac{s^*}{N}}\right)^2 \mid \mathcal{F}_{\frac{i-1}{n}}\right] + \mathrm{o}_p(1).$$

Because $W \perp\!\!\!\perp \eta$ and the points $\left(\frac{s^*}{N}, \frac{t^*}{N}\right)$ follow a uniform distribution on

$$\left\{\left(\frac{s}{N}, \frac{t}{N}\right) : (i-1)\,m \leq j, k \leq im, j \neq k\right\},$$

we deduce that $\left(W_{\frac{t^*}{N}} - W_{\frac{s^*}{N}}\right)$ is still normal distributed with mean zero. The variance of this random variable can be computed easily and is given by

$$\mathbb{E}\left[\left(W_{\frac{t^*}{N}} - W_{\frac{s^*}{N}}\right)^2\right] = \frac{1}{3n} + \mathrm{o}(1).$$

This finishes the proof. ■

A.4. **Proof of Theorem 4**

By the triangle inequality

$$\left|s_{p_{i\Delta,\Delta},m,\alpha_N} - s_{p^*_{i\Delta,\Delta},m}\right| \leq \alpha_N \max_{0\leq s,t\leq m}\left(\left\{\frac{p^*_{\frac{i-1}{n}+\frac{t}{N}}}{\alpha_N}\right\} - \left\{\frac{p^*_{\frac{i-1}{n}+\frac{s}{N}}}{\alpha_N}\right\}\right).$$

Because

$$\left\{\frac{p^*_{\frac{i-1}{n}+\frac{t}{N}}}{\alpha_N}\right\} < 1$$

for any t, we deduce that

$$\left|s_{p_{i\Delta,\Delta},m,\alpha_N} - s_{p^*_{i\Delta,\Delta},m}\right| = \mathrm{O}(\alpha_N) = \mathrm{o}\left(n^{-1/2}\right).$$

Thus,

$$RRV_{RO}^{n,m} = RRV^{n,m} + \mathrm{o}_p(1),$$

whose stochastic limit is given after Equation (**3.9**). ■

References

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2006), Ultra high frequency volatility estimation with dependent microstructure noise, Working paper, Princeton University.

Aldous, D. J. & Eagleson, G. K. (1978), 'On mixing and stability of limit theorems', *Annals of Probability* **6**(2), 325–331.

Alizadeh, S., Brandt, M. W. & Diebold, F. X. (2002), 'Range-based estimation of stochastic volatility models', *Journal of Finance* **57**(3), 1047–1092.

Andersen, T. G., Benzoni, L. & Lund, J. (2002), 'An empirical investigation of continuous-time equity return models', *Journal of Finance* **57**(4), 1239–1284.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2000), 'Great realizations', *Risk* **13**, 105–108.

Back, K. (1991), 'Asset prices for general processes', *Journal of Mathematical Economics* **20**(4), 317–395.

Bandi, F. M. & Russell, J. R. (2005), Microstructure noise, realized variance, and optimal sampling, Working paper, Graduate School of Business, University of Chicago.

Bandi, F. M. & Russell, J. R. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics* **79**(3), 655–692.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M. & Shephard, N. (2006), A central limit theorem for realized power and bipower variations of continuous semimartingales, *in* Y. Kabanov, R. Lipster & J. Stoyanov, eds, 'From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift', Springer-Verlag, pp. 33–68.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2006*a*), Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, Working paper, Nuffield College, University of Oxford.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2006*b*), Subsampling realised kernels, Working paper, Nuffield College, University of Oxford.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society: Series B* **64**(2), 253–280.

Barndorff-Nielsen, O. E. & Shephard, N. (2004*a*), 'Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics', *Econometrica* **72**(3), 885–925.

Barndorff-Nielsen, O. E. & Shephard, N. (2004*b*), 'Power and bipower variation with stochastic volatility and jumps', *Journal of Financial Econometrics* **2**(1), 1–48.

Barndorff-Nielsen, O. E. & Shephard, N. (2006), 'Econometrics of testing for jumps in financial economics using bipower variation', *Journal of Financial Econometrics* **4**(1), 1–30.

Brandt, M. W. & Diebold, F. X. (2006), 'A no-arbitrage approach to range-based estimation of return covariances and correlations', *Journal of Business* **79**(1), 61–73.

Christensen, K. & Podolskij, M. (2006*a*), Range-based estimation of quadratic variation, Working paper, Aarhus School of Business.

Christensen, K. & Podolskij, M. (2006*b*), 'Realized range-based estimation of integrated variance', *Journal of Econometrics* (Forthcoming).

Delattre, S. & Jacod, J. (1997), 'A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors', *Bernoulli* **3**(1), 1–28.

Dijk, D. J. C. V. & Martens, M. (2006), 'Measuring volatility with the realized range', *Journal of Econometrics* (Forthcoming).

Doornik, J. A. (2002), *Object-Oriented Matrix Programming using Ox*, 3 edn, Timberlake Consultants Press.

Fang, Y. (1996), Volatility modeling and estimation of high-frequency data with Gaussian noise, PhD thesis, Sloan School of Management, MIT. Unpublished.

Garman, M. B. & Klass, M. J. (1980), 'On the estimation of security price volatilities from historical data', *Journal of Business* **53**(1), 67–78.

Ghysels, E., Harvey, A. C. & Renault, E. (1996), Stochastic volatility, *in* G. S. Maddala & C. R. Rao, eds, 'Handbook of Statistics: Volume 14', North-Holland, pp. 119–191.

Gloter, A. & Jacod, J. (2001*a*), 'Diffusions with measurement errors. I - local asymptotic normality', *ESAIM: Probability and Statistics* **5**, 225–242.

Gloter, A. & Jacod, J. (2001*b*), 'Diffusions with measurement errors. II - measurement errors', *ESAIM: Probability and Statistics* **5**, 243–260.

Griffin, J. E. & Oomen, R. C. A. (2006), Covariance measurement in the presence of non-synchronous trading and market microstructure noise, Working paper, University of Warwick.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**(2), 127–161.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**(2), 359–379.

Huang, X. & Tauchen, G. (2005), 'The relative contribution of jumps to total price variance', *Journal of Financial Econometrics* **3**(4), 456–499.

Jacod, J. (1994), Limit of random measures associated with the increments of a Brownian semimartingale, Preprint number 120, Laboratoire de Probabilitiés, Université Pierre et Marie Curie, Paris.

Jacod, J. (1997), On continuous conditional Gaussian martingales and stable convergence in law, Seminaire de Probabilities XXXI, 232-246.

Jacod, J. & Protter, P. E. (1998), 'Asymptotic error distributions for the Euler method for stochastic differential equations', *Annals of Probability* **26**(1), 267–307.

Jacod, J. & Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes*, 2 edn, Springer-Verlag.

Jiang, G. J. & Oomen, R. C. A. (2005), A new test for jumps in asset prices, Working paper, Warwick Business School.

Kalnina, I. & Linton, O. (2006), Estimating quadratic variation consistently in the presence of correlated measurement error, Working paper, London School of Economics.

Large, J. (2006), Estimating quadratic variation when quoted prices change by a constant increment, Working paper, All Souls College, University of Oxford.

Oomen, R. C. A. (2005), 'Properties of bias-corrected realized variance under alternative sampling schemes', *Journal of Financial Econometrics* **4**(3), 555–577.

Parkinson, M. (1980), 'The extreme value method for estimating the variance of the rate of return', *Journal of Business* **53**(1), 61–65.

Podolskij, M. & Vetter, M. (2006), Modified bipower variation in the presence of microstructure noise and jumps, Working paper, Ruhr University of Bochum.

Protter, P. E. (2004), *Stochastic Integration and Differential Equations*, 1 edn, Springer-Verlag.

Rényi, A. (1963), 'On stable sequences of events', *Sankhya: The Indian Journal of Statistics; Series A* **25**(3), 293–302.

Rogers, L. C. G. & Satchell, S. E. (1991), 'Estimating variances from high, low, and closing prices', *Annals of Applied Probability* **1**(4), 504–512.

Sheppard, K. (2006), Realized covariance and scrambling, Working paper, University of Oxford.

Zhang, L. (2005), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* (Forthcoming).

Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), 'A tale of two time scales: determining integrated volatility with noisy high-frequency data', *Journal of the American Statistical Association* **100**(472), 1394–1411.

Table 1: Estimators of the IV of INTC.

	Mean	Std.	$\rho(\cdot, RRV_{BC}^{n,m})$	acf(1)	acf(2)	acf(5)	acf(10)
RV^n							
1 tick	16.311	9.186	0.820	0.610	0.522	0.519	0.463
1mn	3.506	1.699	0.817	0.678	0.602	0.570	0.459
2mn	3.205	1.617	0.799	0.590	0.528	0.487	0.396
5mn	2.868	1.683	0.755	0.523	0.440	0.379	0.364
10mn	2.597	1.744	0.684	0.490	0.384	0.283	0.289
15mn	2.572	1.827	0.684	0.433	0.371	0.281	0.317
30mn	2.556	2.099	0.611	0.325	0.322	0.188	0.248
$RRV^{n,m}$							
2mn	4.883	2.181	0.904	0.729	0.672	0.642	0.543
5mn	4.240	1.975	0.883	0.696	0.637	0.587	0.513
$RRV_{BC}^{n,m}$							
2mn	2.311	0.958	1.000	0.456	0.392	0.414	0.352
5mn	2.384	1.084	0.929	0.584	0.509	0.478	0.431
$TSRV(K,J) - aa$							
5mn	2.882	1.667	0.750	0.564	0.464	0.383	0.331

This table presents descriptive statistics for estimators of the IV of INTC. RV^n is the standard realized variance computed at sampling frequencies between 1 tick - 30 minutes, where 1 tick is the RV based on all data. $RRV^{n,m}$ is the RRV of Christensen & Podolskij (2006*b*), $RRV_{BC}^{n,m}$ is the bias-corrected RRV, and $TSRV\ (K,J) - aa$ is the subsampler of Aït-Sahalia et al. (2006). $\rho(\cdot, RRV_{BC}^{n,m})$ is the correlation with $RRV_{BC}^{n,m}$(2mn), which is our preferred range-statistic. acf(r) denotes the rth order autocorrelation.

Table 2: Estimators of the IV of MSFT.

	Mean	Std.	$\rho(\cdot, RRV_{BC}^{n,m})$	acf(1)	acf(2)	acf(5)	acf(10)
RV^n							
1 tick	12.019	7.615	0.788	0.369	0.330	0.327	0.274
1mn	2.189	1.342	0.739	0.699	0.677	0.569	0.484
2mn	1.946	1.297	0.722	0.701	0.652	0.534	0.460
5mn	1.681	1.309	0.696	0.636	0.580	0.491	0.389
10mn	1.484	1.312	0.625	0.671	0.543	0.417	0.350
15mn	1.437	1.334	0.623	0.579	0.492	0.416	0.349
30mn	1.372	1.305	0.589	0.510	0.554	0.436	0.385
$RRV^{n,m}$							
2mn	3.234	1.694	0.874	0.702	0.655	0.582	0.489
5mn	2.759	1.547	0.841	0.707	0.663	0.594	0.484
$RRV_{BC}^{n,m}$							
2mn	1.685	0.853	1.000	0.381	0.328	0.301	0.225
5mn	1.624	0.887	0.920	0.569	0.520	0.476	0.349
$TSRV(K,J) - aa$							
5mn	1.684	1.366	0.686	0.668	0.589	0.467	0.376

This table presents descriptive statistics for estimators of the IV of MSFT. RV^n is the standard realized variance computed at sampling frequencies between 1 tick - 30 minutes, where 1 tick is the RV based on all data. $RRV^{n,m}$ is the RRV of Christensen & Podolskij (2006*b*), $RRV_{BC}^{n,m}$ is the bias-corrected RRV, and $TSRV\ (K,J) - aa$ is the subsampler of Aït-Sahalia et al. (2006). $\rho(\cdot, RRV_{BC}^{n,m})$ is the correlation with $RRV_{BC}^{n,m}$(2mn), which is our preferred range-statistic. acf(r) denotes the rth order autocorrelation.

Table 3: Finite sample properties of asymptotic pivots.

Panel A: $RRV_{BC}^{n,m}$ - Infeasible									
No. obs.	Mean	Std.	0.5%	2.5%	5%	95%	97.5%	99.5%
100	0.346	1.217	0.000	0.003	0.014	0.861	0.899	0.948
200	0.271	1.165	0.000	0.008	0.025	0.880	0.917	0.962
300	0.197	1.117	0.001	0.011	0.030	0.897	0.933	0.972
450	0.138	1.081	0.001	0.014	0.034	0.912	0.945	0.979
600	0.102	1.058	0.001	0.015	0.036	0.921	0.951	0.983
900	0.061	1.038	0.002	0.017	0.040	0.930	0.959	0.987
1200	0.039	1.027	0.002	0.018	0.042	0.934	0.963	0.989
1500	0.031	1.015	0.002	0.018	0.042	0.939	0.966	0.991
Panel B: $RRV_{BC}^{n,m}$ - Feasible								
No. obs.	Mean	Std.	0.5%	2.5%	5%	95%	97.5%	99.5%
100	-0.061	1.168	0.036	0.067	0.093	0.985	0.999	1.000
200	-0.011	1.105	0.026	0.056	0.081	0.969	0.991	1.000
300	-0.023	1.079	0.021	0.051	0.078	0.965	0.989	1.000
450	-0.033	1.056	0.018	0.047	0.074	0.964	0.987	0.999
600	-0.041	1.040	0.016	0.044	0.070	0.964	0.987	0.999
900	-0.051	1.028	0.013	0.041	0.068	0.963	0.985	0.999
1200	-0.056	1.020	0.012	0.039	0.067	0.963	0.985	0.998
1500	-0.052	1.011	0.011	0.037	0.064	0.963	0.985	0.998
Panel C: $\ln RRV_{BC}^{n,m}$ - Feasible								
No. obs.	Mean	Std.	0.5%	2.5%	5%	95%	97.5%	99.5%
100	0.142	1.071	0.007	0.028	0.051	0.922	0.959	0.992
200	0.133	1.068	0.006	0.026	0.050	0.923	0.958	0.990
300	0.093	1.051	0.006	0.026	0.050	0.931	0.962	0.991
450	0.061	1.035	0.005	0.025	0.050	0.938	0.966	0.993
600	0.041	1.023	0.005	0.024	0.049	0.941	0.969	0.993
900	0.018	1.014	0.005	0.024	0.050	0.945	0.971	0.993
1200	0.006	1.008	0.004	0.024	0.049	0.946	0.972	0.994
1500	0.004	1.000	0.004	0.023	0.048	0.948	0.974	0.994

The table shows the finite sample properties of our asymptotic distribution theory. Panel A reports on the convergence in law for $RRV_{BC}^{n,m}$ in the infeasible setting, where $\text{avar}_{RRV_{BC}^{n,m}}$ is known a priori. In Panel B, we estimate $\text{avar}_{RRV_{BC}^{n,m}}$. Panel C is for the feasible log-based distribution theory. The mean, standard deviation and simulated quantiles are shown for the sample sizes $N = 100, 200, 300, 450, 600, 900, 1200, 1500$.

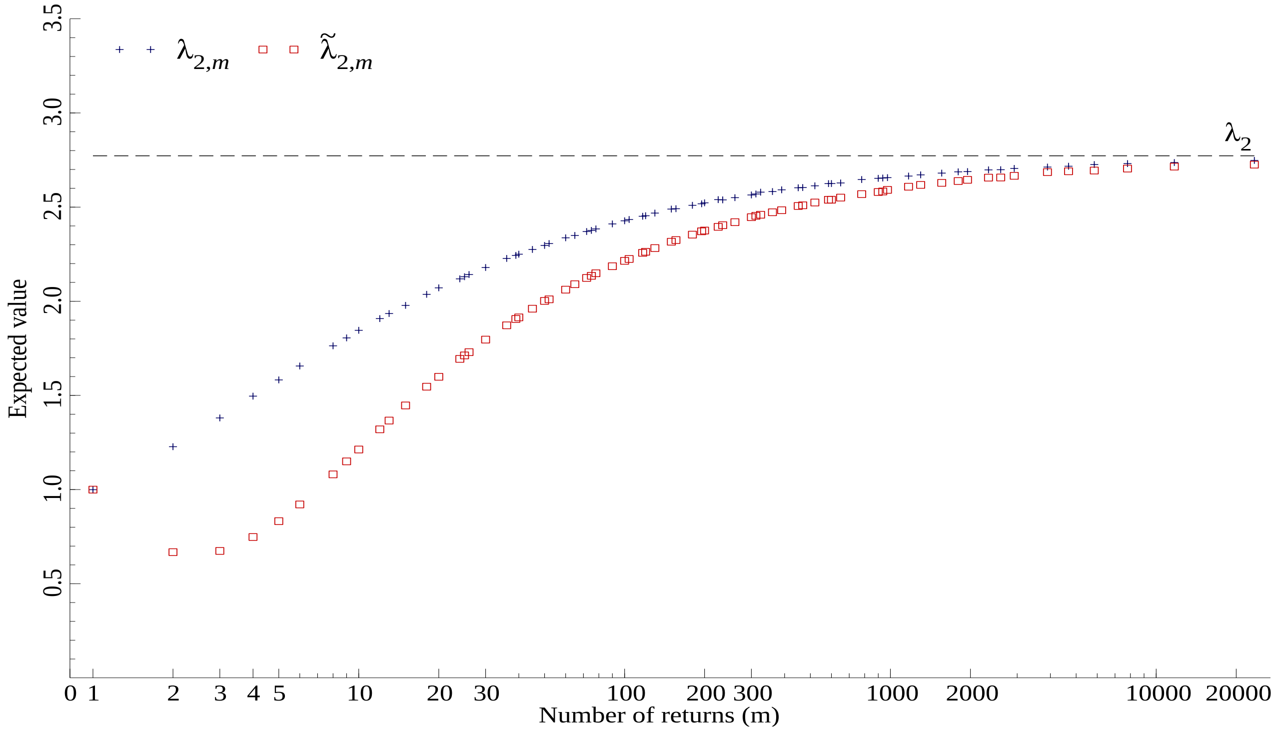

Figure 1: $\lambda_{2,m}$ and $\tilde{\lambda}_{2,m}$ against m on a log-scale. All the estimates are from a simulation with 1,000,000 repetitions. The dashed line is the asymptotic value (as $m \to \infty$).

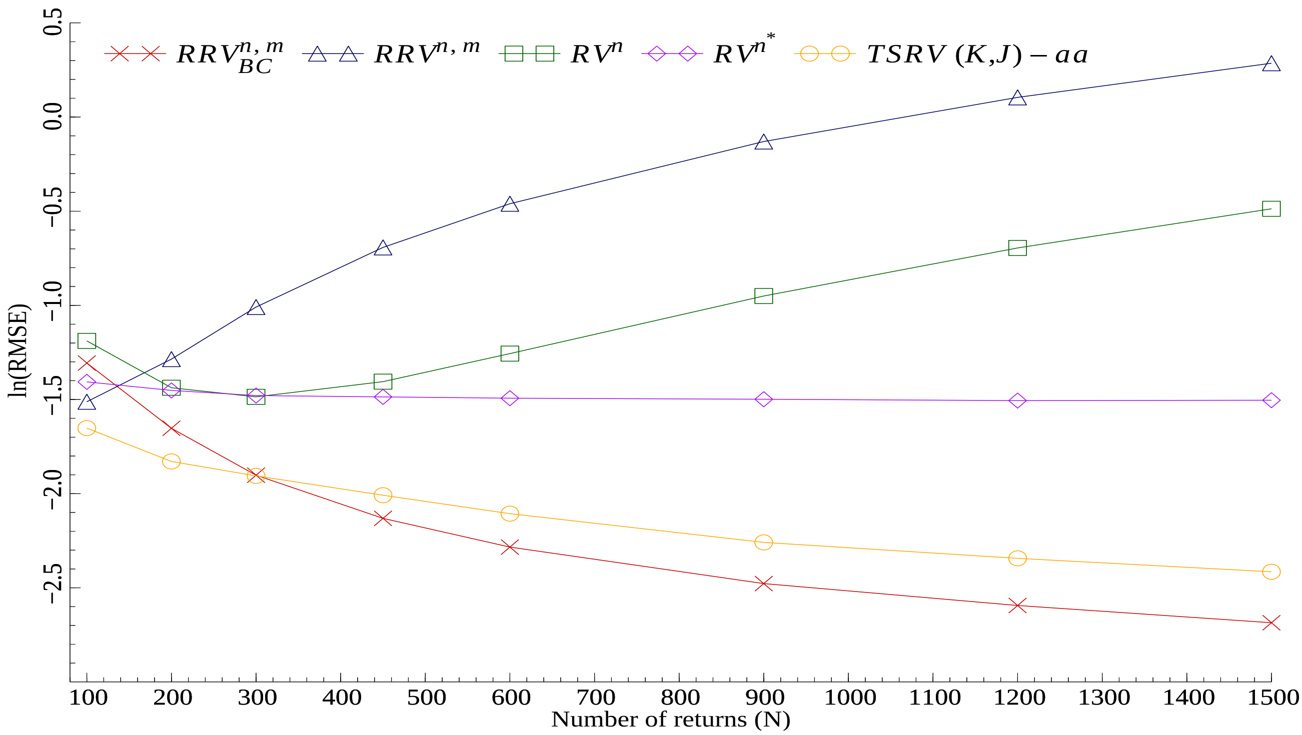

Figure 2: ln(RMSE) of $RRV_{BC}^{n,m}$, $RRV^{n,m}$, RV^n, RV^{n^*} and $TSRV(K,J)-aa$. RV^{n^*} is the RV sampled at an optimal frequency (n^*) using an MSE criterion. $TSRV(K,J)-aa$ is based on an (asymptotically) optimal number of subgrids, where we exhaust the data by shifting the initial point at which prices are recorded. We use the condition $m \geq 10$ to implement $RRV_{BC}^{n,m}$.

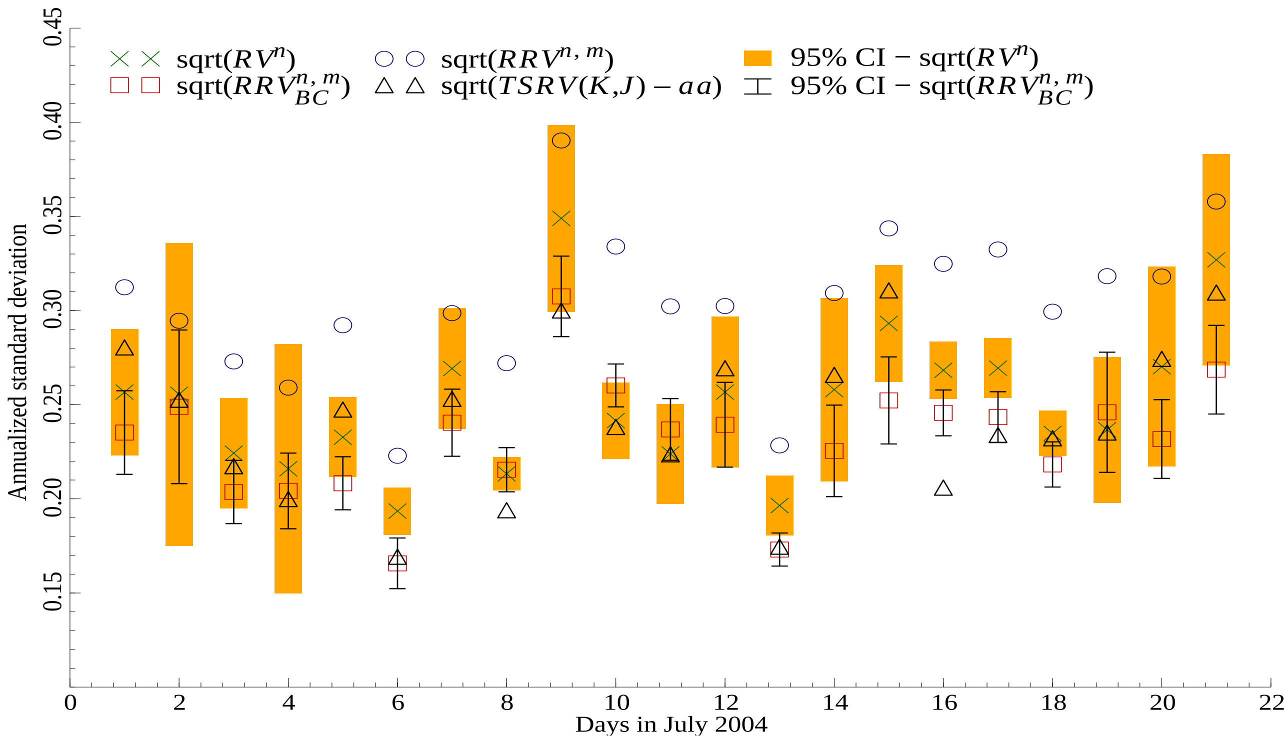

Figure 3: Confidence intervals for the annualized standard deviation of INTC in July, 2004. We plot 95% confidence intervals for $\sqrt{\int_0^1 \sigma_u^2 \mathrm{d}u}$ using the delta method to transform the asymptotic distribution of RV^n and $RRV_{BC}^{n,m}$. The box is based on the feasible limit theory of $\sqrt{RV^n}$ and the line uses that of $\sqrt{RRV_{BC}^{n,m}}$. The point estimates of $\sqrt{RRV^{n,m}}$ and $\sqrt{TSRV\,(K,J) - aa}$ are also reported.

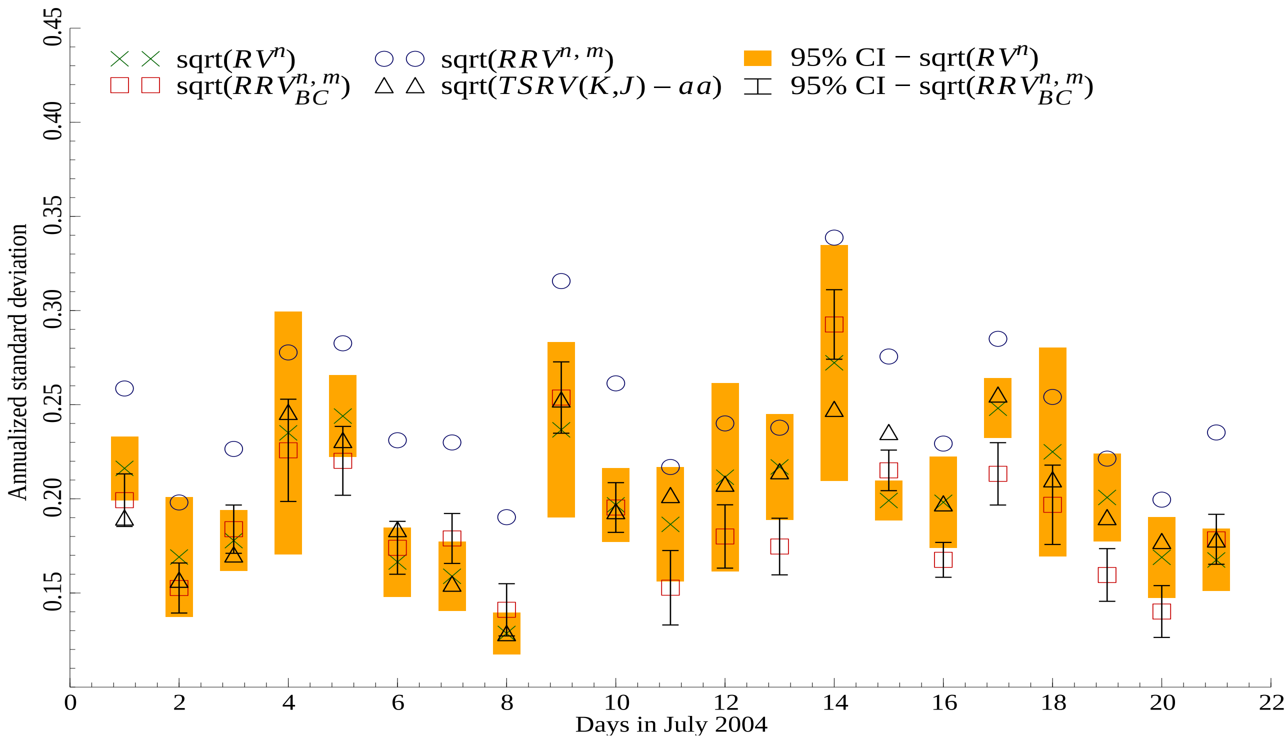

Figure 4: Confidence intervals for the annualized standard deviation of MSFT in July, 2004. We plot 95% confidence intervals for $\sqrt{\int_0^1 \sigma_u^2 \mathrm{d}u}$ using the delta method to transform the asymptotic distribution of RV^n and $RRV_{BC}^{n,m}$. The box is based on the feasible limit theory of $\sqrt{RV^n}$ and the line uses that of $\sqrt{RRV_{BC}^{n,m}}$. The point estimates of $\sqrt{RRV^{n,m}}$ and $\sqrt{TSRV\,(K,J) - aa}$ are also reported.